1, 4

SAMPLE REUSE PROCEDURES FOR PREDICTING THE UNOBSERVED PORTION OF A PARTIALLY OBSERVED VECTOR

by

Seymour Geisser*

Technical Report No. 372

School of Statistics

University of Minnesota

April, 1980

*This work was supported by NIH-GM-25271 research grant.

SAMPLE REUSE PROCEDURES FOR PREDICTING THE UNOBSERVED PORTION OF A PARTIALLY OBSERVED VECTOR

By

Seymour Geisser*
University of Minnesota

1. Introduction

A data analytic aparmetric method termed Predictive Sample Reuse (PSR), Geisser (1974, 1975) is capable of providing solutions to the problem of predicting the unobserved portion of a partially observed vector in the presence of other fully observed vectors of the same kind. In general, PSR methodology assumes a set of N observations or vectors observed at known covariates and a function, which depends on a set of unknown constants, for predicting potential observables generated from the underlying process at known values of the covariates. A set of observables is then deleted and the predictive function, formed on the retained set, is used to predict the deleted observables. A discrepancy function is defined between the predicted values which depend on the unknown constants and the known deleted values for every specified partition of observables into the deleted and retained sets. The discrepancy is then minimized with respect to the set of unknown constants. The solution for the constants is then inserted in the predictive function and is now available for forecasting potential observables.

2. The Prediction Problem

Suppose N units are observed at the same p time points so that

*This work was supported by NIH-GM-25271 research grant.

they are represented by p-dimensional vectors X_α, $\alpha = 1,\ldots,N$. Further, an additional vector X_{N+1} is observed at the first $p_1 < p$ points (in actuality, it may be at any of the p_1 out of p points) and the object is to predict (or retrodict past values) the unobserved p_2 values of this partially observed vector. For convenience, we shall assume that we will be dealing with the last p_2 points and define

$$X_\alpha = \begin{matrix} p_1 \\ p_2 \end{matrix} \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \qquad (1)$$

$\alpha = 1,\cdots,N+1$ and assume $X^{(2)}_{N+1}$ has not been observed and is to be predicted.

3. General PSR Approach - Combination of Predictors

Suppose from the first vectors, $X_1,\cdots,X_N$ each at the same p points we generate a predictor of $X^{(2)}_{N+1}$, say $\hat{X}^{(2)}_{(N)}$. Further, suppose another predictor of $X^{(2)}_{N+1}$ is obtained, say $\hat{X}^{(2)}_{N+1}$, which depends only on the observed $X^{(1)}_{N+1}$. Finally, we combine the two independently calculated predictors into a new predictor

$$\dot{X}^{(2)}_{N+1} = f\left(\hat{X}^{(2)}_{(N)}, \hat{X}^{(2)}_{N+1}; \Omega\right) \qquad (2)$$

for $\Omega \in \mathcal{T}$, $\mathcal{T}$ being the admissible domain of Ω and f an assumed predictive function.

Now let

$$\dot{X}^{(2)}_\alpha = f(\hat{X}^{(2)}_{(N-1,\alpha)}, \hat{X}^{(2)}_\alpha; \Omega) \qquad (3)$$

where $\alpha = 1,\cdots,N$ and $\hat{X}^{(2)}_{(N-1,\alpha)}$ be the predictor for $X^{(2)}_\alpha$ based on $X_1,\cdots,X_{\alpha-1}, X_{\alpha+1},\cdots,X_N$ and of the same functional form as $\hat{X}^{(2)}_{(N)}$ and

$\hat{X}_{\alpha}^{(2)}$ be the predictor of $X_{\alpha}^{(2)}$ based only on $X_{\alpha}^{(1)}$ and of the same functional form as $\hat{X}_{N+1}^{(2)}$. Further, define an overall discrepancy measure

$$D(\Omega) = D(d_1, d_2, \ldots, d_N) , \tag{4}$$

where $d_{\alpha} = d\left(\dot{X}_{\alpha}^{(2)}, X_{\alpha}^{(2)}\right)$ is some measure of the discrepancy of $\dot{X}_{\alpha}^{(2)}$ and the actual value it is predicting, namely $X_{\alpha}^{(2)}$. This discrepancy measure is then minimized with respect to Ω within its given domain of definition T. If $\hat{\Omega}$ is the unique solution which satisfies the constraints, then the final predictor is given as

$$\tilde{X}_{N+1}^{(2)} = f\left(\hat{X}_{(N)}^{(2)}, \hat{X}_{N+1}^{(2)}; \hat{\Omega}\right) . \tag{5}$$

An interesting special case for the predictive function is

$$\dot{X}_{N+1}^{(2)} = \Omega \hat{X}_{(N)}^{(2)} + (I - \Omega) \hat{X}_{N+1}^{(2)} \tag{6}$$

where Ω is real $p_2 \times p_2$ matrix such that Ω and $I - \Omega$ possess only non-negative roots. Define

$$\dot{X}_{\alpha}^{(2)} = \Omega \hat{X}_{(N-1,\alpha)}^{(2)} + (I - \Omega) \hat{X}_{\alpha}^{(2)} \tag{7}$$

and assume a simple quadric at discrepancy

$$D(\Omega) = \sum_{\alpha=1}^{N} \left(\dot{X}_{\alpha}^{(2)} - X_{\alpha}^{(2)}\right)' \left(\dot{X}_{\alpha}^{(2)} - X_{\alpha}^{(2)}\right) . \tag{8}$$

If this combination of predictive function and discrepancy measure is deemed realistic then a solution may easily be obtained for general forms of $\hat{X}_{(N)}^{(2)}$ and $\hat{X}_{N+1}^{(2)}$ since (8) is easily shown to be equivalent to

$$D(\Omega) = \mathrm{Tr}[(\Omega - \hat{\Omega})B(\Omega - \hat{\Omega})' + F]$$

where F does not depend on Ω and

$$\hat{\Omega} = \sum_{\alpha=1}^{N} \left(X_{\alpha}^{(2)} - \hat{X}_{\alpha}^{(2)}\right)\left(\hat{X}_{(N-1,\alpha)}^{(2)} - \hat{X}_{\alpha}^{(2)}\right)\left[\sum_{\alpha=1}^{N} \left(\hat{X}_{(N-1,\alpha)}^{(2)} - \hat{X}_{\alpha}^{(2)}\right)\left(\hat{X}_{(N-1,\alpha)}^{(2)} - \hat{X}_{\alpha}^{(2)}\right)'\right]^{-1}$$

$$= AB^{-1} \tag{9}$$

Hence

$$\tilde{X}_{N+1}^{(2)} = \hat{\Omega}\,\hat{X}_{(N)}^{(2)} + (I-\hat{\Omega})\,\hat{X}_{N+1}^{(2)} \tag{10}$$

$$= \hat{X}_{(N)}^{(2)} + (I - \hat{\Omega})\left(\hat{X}_{N+1}^{(2)} - \hat{X}_{(N)}^{(2)}\right)$$

provided $\hat{\Omega}$ exists and satisfies whatever constraints on $\hat{\Omega}$ deemed appropriate. A particular constraint which requires the roots of Ω to lie in $[0,1]$ can easily be imposed, if the solution $\hat{\Omega}$ has a reduced characteristic function posessing only distinct linear factors. Clearly we may write

$$\hat{\Omega} = P\,D\,P^{-1}$$

where P is a nonsingular matrix and D is the diagonal matrix with the roots of $\hat{\Omega}$ along the main diagonal. Now replace any root outside $[0,1]$ by 0 or 1 depending on which is closer and denote the new diagonal matrix as D_1. The constrained solution is then

$$\hat{\Omega}_c = P\,D_1\,P^{-1} \quad . \tag{11}$$

and yields what we shall refer to as the Combined (C) predictor.

Note that if the discrepancy of (8) is modified by a known $p_2 \times p_2$ p.d. matrix Λ so that

$$D(\Omega) = \sum_{=1}^{N} (\dot{X}_{\alpha}^{(2)} - X_{\alpha}^{(2)})'\Lambda(\dot{X}_{\alpha} - X_{\alpha}^{(2)})$$

then the solution (9) is simply modified by the symmetric square root of Λ to

$$\hat{\Omega} = \Lambda^{\frac{1}{2}} AB^{-1} \Lambda^{-\frac{1}{2}} \tag{12}$$

4. A Special Case and Regression Predictors

A further special case useful in simple regression or growth curve situations is where the fitted equations are of the form

$$X = Z_{pxm}\, B_{mx1} = \begin{matrix} p_1 \\ p_2 \end{matrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} B \tag{13}$$

for known Z and $B = (Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}X$ where $\bar{X} = N^{-1} \sum_{\alpha=1}^{N} X_\alpha$, $m \leq p_1, p_2$ arbitrary and Σ is a p.d. matrix partitioned as follows

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{22} & \Sigma_{21} \end{pmatrix} .$$

Here we may take

$$\hat{X}^{(2)}_{(N)} = Z_2\, B \tag{14}$$

and

$$X^{(2)}_{N+1} = Z_2'(Z_1'\Sigma_{11}^{-1}Z_1)^{-1}Z_1'\Sigma_{11}^{-1}X^{(1)}_{N+1} \ . \tag{15}$$

A particular solution was given by Geisser (1975) for $m = 2$ and $p_2 = 1$, $\Sigma = I$ and Ω a 1×1 scalar. There

$$Z' = \begin{pmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_p \end{pmatrix} \tag{16}$$

so that one is calculating a convex combination of two predicted values at z_p, one computed from the average of the fitted straight lines of the first N vectors and the other is the predicted value obtained from the line fitted to the $N+1$ vector using the first $p-1$ observed points. Substitution in (9) yields $\hat{\omega}$, the mixing constant, as a very special case for the more general vector solution (10).

However, another predictive function which may be pertinant for the situation described by (13) is a PSR simulation of the conditional expectation predictor as typically occurs for multivariate normal populations namely

$$\dot{X}^{(2)}_{N+1} = \hat{X}^{(2)}_{(N)} + \Omega\left(X^{(1)}_{N+1} - \hat{X}^{(1)}_{(N)}\right) \tag{17}$$

where Ω is a $p_2 \times p_1$ "regression" matrix and

$$\hat{X}^{(i)}_{(N)} = Z_i B, \quad i = 1, 2 . \tag{18}$$

$$\dot{X}^{(2)}_{\alpha} = \hat{X}^{(2)}_{(N-1,\alpha)} + \Omega\left(X^{(1)}_{\alpha} - \hat{X}^{(1)}_{(N-1,\alpha)}\right) . \tag{19}$$

Then for a simple quadratic discrepancy measure,

$$\sum_{\alpha=1}^{N} \left(\dot{X}^{(2)}_{\alpha} - X^{(2)}_{\alpha}\right)'\left(\dot{X}^{(2)}_{\alpha} - X^{(2)}_{\alpha}\right) = \mathrm{Tr}[(\Omega-\hat{\Omega})E(\Omega-\hat{\Omega})' + G] \tag{20}$$

where G does not depend on Ω, unconstrained minimization with respect to Ω yields

$$\hat{\Omega} = \sum_{\alpha}\left(X^{(2)}_{\alpha}-\hat{X}^{(2)}_{(N-1,\alpha)}\right)\left(X^{(1)}_{\alpha}-\hat{X}^{(1)}_{(N-1,\alpha)}\right)'\left[\sum_{\alpha}\left(X^{(1)}_{\alpha}-\hat{X}^{(1)}_{(N-1,\alpha)}\right)\left(X^{(1)}_{\alpha}-\hat{X}^{(1)}_{(N-1,\alpha)}\right)'\right]^{-1}$$
$$= CE^{-1} \tag{21}$$

and Unconstrained Regression (UR) predictor

$$\tilde{X}^{(2)}_{N+1} = \hat{X}^{(2)}_{(N)} + \hat{\Omega}\left(X^{(1)}_{N+1} - \hat{X}^{(1)}_{(N)}\right) . \tag{22}$$

The insinuation of a known p.d. matrix Λ in (20) results in modifying (21) to $\Lambda^{\frac{1}{2}}\hat{\Omega}$.

We note this regression type predictor requires that $m \leq p$ and $p_1 \leq N$ but does not require $m \leq p_1$ as did (10) which was a combination of two independently calculated predictors. Of course in any situation where both are computable both could be examined. It is anticipated that the predictor specified by (22) will be superior when the variation about each individual predictive function is considerable while the variation between their predictive functions is small. On the other hand when the variation about individual vectors is small but considerable variation exists between individual predictive functions, then the predictor (10) should be superior. Certain constraints may also be imposed on Ω when it is thought appropriate. In particular constraining all of the columns but the last to be zero while that column's components are decreasing in absolute

magnitude would mimic to some extent a first order Markov model. Specifically we could assume a Markovian form for the regression matrix

$$\Omega_M = (\underset{\sim}{0}, \underset{\sim}{0}, \ldots, \underset{\sim}{0}, \omega) \tag{23}$$

where $\underset{\sim}{0}$ is a $p_2 \times 1$ null vector and

$$\omega = \begin{pmatrix} \omega_1 \\ \cdot \\ \cdot \\ \omega_{p_2} \end{pmatrix} . \tag{24}$$

If the values of ω are unconstrained, the solution for the simple quadratic discrepancy of equation (20), which reduces to

$$\sum_\alpha (y_\alpha \omega - Z_\alpha)'(y_\alpha \omega - Z_\alpha), = [(\omega - \hat{\omega})'(\omega - \hat{\omega}) + K] \sum_\alpha y_\alpha^2 \tag{25}$$

where K does not depend on ω, is

$$\hat{\omega} = \frac{\sum_\alpha y_\alpha Z_\alpha}{\sum_\alpha y_\alpha^2} = \frac{\sum_\alpha (x_{p_1\alpha} - \hat{x}_{(N-1,\, p_1,\alpha)})\,(X_\alpha^{(2)} - \hat{X}^{(2)}_{(N-1,\alpha)})}{\sum_\alpha (x_{p_1\alpha} - \hat{x}_{(N-1,p_1,\alpha)})^2} \tag{26}$$

where $x_{p_1\alpha}$ is the last component of $X_\alpha^{(1)}$ and $\hat{x}_{(N-1,p_1,\alpha)}$ is the last component of $\hat{X}^{(1)}_{(N-1,\alpha)}$ with obvious definitions for Z_α and y_α. The resulting predictor is denoted as a Markovian Regression (MR) predictor.

To obtain a solution subject to constraints

$$|\omega_1| \geq |\omega_2| \geq \cdots \geq |\omega_{p_2}| \geq 0 \tag{27}$$

proceed as follows: Suppose $\hat{\omega}_k$ has the largest absolute value. Now if $\bar{\omega}_k = k^{-1} \sum_{j=1}^{k} |\omega_j|$ is larger than every $|\hat{\omega}_j|$ for $j > k$, then the solution is $\tilde{\omega}_k = \bar{\omega}_k \operatorname{sgn} \hat{\omega}_j$ $j = 1, \ldots, k$ where for any real

$r \neq 0$, sgn $r = |r|\ r^{-1}$. If this is not the case, continue sequentially until the average $\bar{\omega}_{k+t} = (k+t)^{-1} \sum_{j=1}^{k+t} |\hat{\omega}_j|$ exceeds the absolute value of subsequent $\hat{\omega}_{k+t+1}, \ldots, \hat{\omega}_{p_2}$ for the first time. Then $\tilde{\omega}_j = \bar{\omega}_{k+t}$ sgn $\hat{\omega}_j$ for $j = 1, \ldots, k + t$, and restart the process with the reduced subset $(\hat{\omega}_{k+t+1}, \ldots, \hat{\omega}_{p_2})$. The process will continue until the last average say $\bar{\omega}_m = m^{-1} \sum_{i=s+1}^{s+m} |\hat{\omega}_i|$ exceeds in absolute value all subsequent $\hat{\omega}_i$ and $|\hat{\omega}_i| \geq |\hat{\omega}_{i+1}|$ for all remaining i. These $\hat{\omega}_i$'s are then retained in the solution. In case at any point the $\max_i |\omega_i|$ is not unique it will always speed up the process to choose the largest index k for $\max_i |\omega_i| = |\omega_k|$. Whenever a component $\hat{\omega}_j = 0$ and a non-zero component $\hat{\omega}_k$ exists such that $k > j$ then multiple solutions will ensue, since $\tilde{\omega}_k$ admits as solutions both $\pm \bar{\omega}_{k+t}$. The resulting predictor is denoted as a Decreasing Markovian Regression (DMR) predictor. To obtain an Increasing Markovian Regression (IMR) predictor we reverse the constraint i.e.

$$0 \leq |\omega_1| \leq |\omega_2| \leq \ldots |\omega_{p_2}| \quad .$$

The solution is derived in exactly the same fashion as for (DMR) except the process is initiated from the opposite end.

Sometimes a constraint in addition to the one given by (27), appears sensible. Namely we may require

$$1 \geq |\omega_1| \geq |\omega_2| \ldots\ldots \geq |\omega_{p_2}| \geq 0 \tag{28}$$

One covariance structure which satisfies (28) is such that $\text{corr}(X_i, X_j) = \rho^{|i-j|}$ and $\text{Var}(X_i) = \theta_i^2$ with the additional restriction that

$$\theta_i \geq |\rho| \theta_{i+1} \quad . \tag{29}$$

The regression matrix Ω_M of (23) would then have as elements

$$\omega_i = \rho^i \theta_{p_i+i} / \theta_{p_1} \qquad i = 1, \ldots, p_2 \; . \tag{30}$$

Clearly such a model while retaining greater latitude then the conventional serial model where $\theta_i^2 \equiv \theta^2$ is still less general than that specified by the constraints of (28).

To obtain the solution subject to the constraint specified by (28), one starts with the solution to (27), say $\tilde{\omega}_i$, and if it does not conform to the final constraint the new solution is

$$\dot{\omega}_1 = \text{sgn } \tilde{\omega}_1$$

and all subsequent $\tilde{\omega}_i$ are obtained by dividing by $|\tilde{\omega}_1|$ until the first index k such that $|\dot{\omega}_k| > |\tilde{\omega}_{k+1}|$. The final solution then is $\dot{\omega}_i = \tilde{\omega}_i |\tilde{\omega}_1|^{-1}$, for $i = 2, \ldots, k$ and $\dot{\omega}_i = \tilde{\omega}_i$ for $i = k+1, \ldots p_2$. This yields a Constrained Decreasing Markovian Regression (CDMR) predictor.

5. Illustrations:

We illustrate these techniques on two sets of data. The first set of data on 20 boys whose ramus height was linearly regressed on age, for the years 8, 8½, 9, 9½, is from Elston and Grizzle (1962, Table 2). This data was used by Lee and Geisser (1975) to illustrate the computation of 16 different Bayesian, frequentist and heuristic predictors developed by Geisser (1970) and Lee and Geisser (1972). Fearn (1975) also used this data to demonstrate a particular Bayesian predictor along the lines of Lindley and Smith (1972). We obtain the predicted values here by setting $\Sigma = I$. As previously, the last observed value at age 9½ in one vector (boy) is

withheld and the rest of the data is used to predict that value for each of the various methods. This is repeated in turn for each vector and the average absolute and squared deviations are computed and compared in Table 1. The predictor which is clearly best for both the mean squared and absolute deviation is the CDMR predictor. This is not surprising as the best of those considered by Lee and Geisser (1975) was a first order serial correlation model whose mean squared deviation was .5096 as compared to the CDMR predictor which was .5081. There then is little to distringuish between these two. The Fearn-Lindley Bayesian predictor yields a mean squared deviation which is 10% larger than the CDMR predictor and about 4% larger than the MR predictor and is, by the way, considerably more difficult to compute than either. For this data set, the constraint required that $\dot{\omega}_1 = 1$ for the CDMR predictor of the last observation. Hence for its calculation one merely added to the usual regression predictor from the other vectors the difference between the last observed value of the designated vector and the usual regression predictor of that value from the other vectors.

Quadratic predictors were also calcualted to determine if any improvement could be made. The quadratic predictors increased the mean squared deviations for each predictor - about 5-6% for all but C where the new mean squared deviation was almost four times as large as for the linear predictor. It would appear that C is very sensitive to the general form of the predictive function while the others tend to adjust the function more to the data at hand.

Table 1: Observed and predicted ramus heights of final measurements in mm.

Boy	Observed Value	Linear Predictors C	UR	MR	CDMR
1	49.7	49.888	49.768	49.942	49.951
2	48.4	48.714	48.563	48.637	48.648
3	48.5	48.739	48.863	48.730	48.742
4	47.2	46.867	47.102	46.963	47.045
5	49.3	49.805	49.743	49.851	49.853
6	53.7	53.591	53.997	54.320	54.263
7	54.5	55.695	55.358	55.342	55.223
8	52.7	50.599	50.944	51.233	51.235
9	54.4	54.318	52.362	53.157	53.157
10	48.3	49.006	47.968	48.188	48.228
11	51.9	51.771	52.413	52.584	52.570
12	55.5	54.357	53.867	53.788	53.788
13	55.0	54.147	54.501	54.623	54.623
14	49.8	50.041	50.138	50.250	50.255
15	51.8	51.955	52.095	52.149	52.143
16	53.3	53.512	53.614	53.665	53.636
17	49.5	49.168	49.325	49.316	49.342
18	55.3	55.790	55.931	56.230	56.048
19	48.4	49.404	49.036	49.053	49.053
20	51.8	53.270	53.708	52.183	52.176
Mean abs. dev.		.5900	.6679	.5934	.5672
Mean squ. dev.		.6304	.8248	.5401	.5081

C: Combination of predictors

UR: Unconstrained Regression predictor

MR: Markovian Regression predictor

*IMR: Increasing Markovian Regression predictor

*DMR: Decreasing Markovian Regression predictor

CDMR: Constrained Decreasing Markovian Regression predictor

*In this Table IMR and DMR are equivalent to the MR predictor because only one value is being predicted.

Table 1a: Observed and predicted ramus heights of final measurements in mm.

Boy	Observed Value	Quadratic Predictors C	UR	MR*	CDMR
1	49.7	49.453	49.721	49.906	49.912
2	48.4	48.991	48.521	48.602	48.606
3	48.5	50.111	48.841	48.684	48.689
4	47.2	49.031	47.072	46.897	46.974
5	49.3	49.746	49.709	49.820	49.820
6	53.7	52.294	53.967	54.274	54.223
7	54.5	53.773	55.307	55.332	55.218
8	52.7	50.930	50.830	51.127	51.130
9	54.4	52.147	52.270	53.122	53.122
10	48.3	47.767	47.869	48.177	48.207
11	51.9	51.675	52.403	52.532	52.520
12	55.5	57.538	53.737	53.687	53.687
13	55.0	53.537	54.451	54.551	54.551
14.	49.8	50.128	50.108	50.210	50.213
15	51.8	51.809	52.063	52.106	52.101
16	53.3	52.900	53.578	53.623	53.599
17	49.5	50.072	49.286	49.256	49.279
18	55.3	53.515	55.890	56.205	56.030
19	48.4	49.262	48.993	40.039	49.039
20	51.8	55.923	53.702	52.155	52.149
Mean abs. dev.		1.1610	.6744	.5935	.5685
Mean squ. dev.		2.2712	.8761	.5646	.5339

*In this table IMR and DMR are equivalent to the MR predictor because only one value is being predicted.

A second set of data, Table 2, was abstracted from Bliss (1970). Here the weight in grams in 10 day intervals is recorded for 15 male rats. Although the data were recorded for 210 days, we have only used the data up to day 90 where a linear fit was, at first glance, not unreasonable and we set $\Sigma = I$.

Table 2: Weight in grams of rats on stock diet at 50 to 90 days of age.

	Age				
Rat	50	60	70	80	90
1	210	274	318	360	404
2	222	278	346	385	430
3	180	200	260	290	319
4	182	194	250	296	340
5	178	202	240	278	306
6	170	207	244	275	308
7	188	240	280	318	354
8	200	244	292	340	392
9	185	238	282	332	362
10	211	276	298	325	367
11	204	268	340	392	430
12	224	314	362	408	436
13	177	222	286	321	365
14	206	224	260	290	304
15	183	239	274	318	342

Here we were interested in predicting the last two measurements i.e. for day 80 and day 90. As before these two values were withheld and predicted from the rest of the data for each rat in turn and the mean squared deviation computed for both these days. The observed and predicted values are recorded in Table 3. Note that again the various Markov regression methods MR, DMR and CDMR are clearly superior overall to C and UR. The CDMR predictor is best for day 80 while the DMR is best for day 90, but the MR is quite close to both of them. A word of caution concerning this data must be mentioned here. A scan of the entire average curve of this data, i.e. through day 210 Bliss (1970, Fig. 4.2), indicates that the data are well fitted by an exponential curve with an upper asymptote and the slight but perceptible bend away from linearity begins at about day 80. Hence most of the linear predictors tend to overshoot at day 80 and do so with greater frequency at day 90. Hence, quadratic predictors

may produce a better fit for the five data points. The computation for the quadratic predictors is given in Table 3a. Clearly, the quadratic fit substantially improved the predicted values for days 80 and 90 over the linear fit for every predictor and most emphatically on day 90. The predictor that made the most dramatic relative gain was C. Overall it was now just behind the leader DMR with MR closely following it and CDMR next. By an appreciabe amount, UR was still the worst. Throughout, the primary component is the fit of the approximately assumed function, and secondarily the covariational relationships. Further, C is most sensitive to the fitted predictive function or to put it another way, perhaps least robust.

6. Remarks:

Although each of the methods as presented yields only a point predictor it is possible to obtain a partial ordering of predictive values using $D(\Omega)$. If $D(\Omega_0) < D(\Omega_1)$ we assert that $\dot{X}(\Omega_0)$ is a less discrepant predictor than $\dot{X}(\Omega_1)$. Ordering values in this manner would result in a least discrepant β-set of predictors attained by calculating all values of $\dot{X}(\Omega)$ which satisfy

$$\frac{D(\hat{\Omega})}{D(\Omega)} \geq \beta$$

Hence $\tilde{X} = \dot{X}(\hat{\Omega})$, the point predictor we previously used, now becomes the least discrepant predictor (assuming it is unique). Of course this is a less informative ordering than induced by posterior probabilities or likelihoods, but may adequately serve in the presence of a less informative structure.

I am indebted to Dennis Jennings for programming and executing the calculations herein.

Table 3: Observed and predicted weights in grams for days 80 and 90 for rats.

Rat	Observed Values at 80 and 90 Days	Linear Predictors C	UR	MR*	DMR	CDMR
1	360	364.53	363.09	363.27	364.23	361.12
	404	408.66	406.52	407.43	406.46	403.35
2	385	387.91	397.75	394.71	396.23	389.25
	430	432.24	445.73	439.63	438.11	431.13
3	290	298.78	305.64	301.11	300.35	302.28
	319	339.16	350.79	342.57	343.33	345.26
4	296	284.26	289.80	289.11	287.65	291.30
	340	322.21	331.27	328.82	330.28	333.93
5	278	280.62	279.48	279.33	278.21	282.21
	306	321.34	319.55	320.33	321.45	325.45
6	275	290.32	285.83	284.23	282.91	285.92
	308	329.66	322.00	324.75	326.07	329.08
7	318	326.41	322.89	322.36	322.23	322.57
	354	368.27	362.85	364.79	364.92	364.25
8	340	334.19	335.25	334.92	335.18	334.43
	392	376.50	378.10	377.62	377.36	376.61
9	332	328.01	323.76	324.27	324.22	324.36
	362	369.76	363.92	366.58	366.63	366.78
10	325	348.14	345.10	342.42	342.87	341.44
	367	393.32	387.77	386.34	385.90	384.48
11	392	382.65	386.40	384.99	387.07	383.46
	430	426.98	433.19	430.64	428.57	424.96
12	408	414.17	410.68	410.76	416.23	408.96
	436	467.53	463.76	463.61	458.14	450.88
13	321	332.95	334.07	328.38	328.44	328.28
	365	369.48	371.21	370.73	370.67	370.50
14	290	298.38	304.98	302.02	301.59	303.34
	304	349.37	362.83	344.90	345.33	347.08
15	318	322.13	315.24	315.91	315.61	316.50
	342	364.18	356.38	358.09	358.39	359.28
Mean squ. dev. 80		99.786	102.293	68.736	72.789	67.782
Mean squ. dev. 90		414.943	460.879	314.400	302.464	311.619

*In this Table MR and IMR are equivalent.

Table 3a: Observed and predicted weights in grams for days 80 and 90 for rats.

Rat	Observed Values at 80 and 90 Days	Quadratic Predictors C	UR	MR*	DMR	CDMR
1	360	359.68	359.27	360.97	361.99	358.13
	404	396.85	395.58	398.93	397.91	394.05
2	385	394.28	394.06	392.39	394.08	385.82
	430	437.19	434.30	431.48	429.79	421.53
3	290	302.90	302.95	298.41	297.39	300.73
	319	342.86	342.99	332.59	333.61	336.95
4	296	284.88	284.00	285.58	283.52	290.04
	340	317.89	315.68	316.31	318.37	324.89
5	278	275.62	275.44	276.11	274.48	280.89
	306	308.60	308.13	309.03	310.65	317.06
6	275	282.65	284.10	281.40	279.64	284.91
	308	314.52	317.41	314.30	316.06	321.34
7	318	320.05	320.42	319.83	319.52	320.54
	354	354.72	355.70	355.48	355.78	356.80
8	340	331.62	330.96	332.17	332.33	331.79
	392	367.73	365.87	367.48	367.32	366.78
9	332	322.96	323.11	321.77	321.55	322.30
	362	358.20	359.78	357.42	357.63	358.38
10	325	338.23	338.55	339.98	340.37	338.93
	367	369.97	366.09	377.48	377.08	375.64
11	392	386.27	386.66	383.04	385.17	379.55
	430	431.22	433.40	422.95	420.82	415.20
12	408	408.16	408.02	409.34	415.10	402.24
	436	455.69	455.33	357.76	451.00	439.14
13	321	331.33	334.93	325.85	325.76	326.06
	365	369.44	374.24	361.42	361.51	361.00
14	290	294.76	294.36	299.27	298.62	301.50
	304	331.27	342.45	334.97	335.62	338.51
15	318	313.63	314.11	313.44	312.97	314.74
	342	347.95	350.45	349.17	349.64	351.41
Mean squ. dev. 80		63.539	72.699	59.014	63.760	65.336
Mean squ. dev. 90		200.707	270.750	206.896	196.576	218.472

*In this Table MR and IMR are equivalent.

References

[1] Bliss, C.I. (1970 Statistics in Biology V.2, New York, McGraw-Hill.

[2] Elston, R.C. & Grizzle, J.E. (1962). Estimation of time-response curves and their confidence bands. Biometrics 18, 148-59.

[3] Fearn, T. (1975). A Bayesian approach to growth curves. Biometrika, 62, 1, 89-100.

[4] Geisser, S. (1970). Bayesian analysis of growth curves. Sankhyā A 32, 53-64.

[5] Geisser, S. (1974), A predictive approach to the random effect model, Biometrika, 61, 101-107.

[6] Geisser, S. (1975), The predictive sample reuse method with applications, Journal of the American Statistical Association, 70, 320-328.

[7] Lee, J.C. & Geisser, S. (1972). Growth curve prediction. Sankhyā A 34, 393-412.

[8] Lee, J.C. & Geisser, S. (1975). Application of growth curve prediction. Sankhyā A 37, 239-256.

[9] Lindley, D.V. & Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion), J.R. Statist. Soc. B34, 1-41.